Decision Support Systems 50 (2011) 570–575

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Detecting evolutionary financial statement fraud

Wei Zhou [a,*], Gaurav Kapoor [b]

[a] Information Systems and Technologies, ESCP Europe, 75543 Paris cedex 11, France
[b] Information Systems and Operations Management, University of Florida, Gainesville, Florida 32611, USA

ARTICLE INFO

Available online 24 August 2010

Keywords:
Financial statement fraud
Data mining technique
Neural networks

ABSTRACT

A fraudulent financial statement involves the intentional furnishing and/or publishing of false information in it and this has become a severe economic and social problem. We consider Data Mining (DM) based financial fraud detection techniques (such as regression, decision tree, neural networks and Bayesian networks) that help identify fraud. The effectiveness of these DM methods (and their limitations) is examined, especially when new schemes of financial statement fraud adapt to the detection techniques. We then explore a self-adaptive framework (based on a response surface model) with domain knowledge to detect financial statement fraud. We conclude by suggesting that, in an era with evolutionary financial frauds, computer assisted automated fraud detection mechanisms will be more effective and efficient with specialized domain knowledge.

1. Introduction

Since the booming of the Internet and the invention of other modern technologies, there has been a dramatic increase in fraudulent schemes associated with all facets in the business world. Some of these commonly observed schemes include credit card fraud, financial statement fraud, e-commerce transaction fraud, insurance fraud, money laundering, computer intrusion fraud, telecommunications fraud, and subscription fraud. Statistic and machine learning based technologies have been shown to be an effective way to deter and detect fraud, but fraudsters are adaptive and are usually able to find ways to circumvent them. Existing fraud detection techniques for most of the situations involving fraud usually share very similar data mining principles, but they can differ in many aspects with specialized domain knowledge [5].

Financial statement fraud in particular has cast rapidly increasing adverse impact not only on individual investors but the overall stability of global economies. Although there are minor variations in its definition, a financial statement fraud is defined by the Association of Certified Fraud Examiners as “The intentional, deliberate, misstatement or omission of material facts, or accounting data which is misleading and, when considered with all the information made available, would cause the reader to change or alter his or her judgment or decision.” In practice, financial statement fraud might involve: (1) manipulation of financial records, (2) intentional omission of events, transactions, accounts, or other significant information from which financial statements are prepared, or (3) misapplication of accounting principles, policies, and procedures used to measure, recognize, report, and disclose business transactions [20].

Many techniques based on data mining have been investigated and implemented to detect financial statement fraud, including regression, decision trees, neural networks and Bayesian belief networks [12]. These techniques have been shown to be successful in their early stages. However, there is no agreement on which data features and techniques are best for detection. Also, while supervised learning techniques have been among the dominant methods used for detecting financial statement fraud, a majority of related implementations do not keep track of new variations in the methods designed for committing fraud. Moreover, financial fraud is becoming more and more difficult to detect using the current detection techniques. A CEO who is truly knowledgeable and wants to really commit a crime has the necessary resources to easily outwit the system and is able to fool any detection mechanism [7].

Despite the increased of time and effort that has been spent to detect the same, the number of detected frauds[1] and the detection rate[2] have largely decreased [7]. When the executives who are involved in financial fraud are well aware of the fraud detection techniques and software, which are usually public information and are easy to obtain, they are likely to adapt the methods in which they commit fraud and make it difficult to detect the same, especially by existing techniques. There exists an urgent need for new methods that is not only efficient but effective to catch up with these probable newly emerged or adaptive financial shenanigans. We (1) consider existing detection techniques based on data mining, (2) provide an

* Corresponding author.
E-mail addresses: wzhou@escpeurope.eu (W. Zhou), grkapoor@ufl.edu (G. Kapoor).

[1] In 2003, the SEC issued 77 financial statement fraud AAERs (Accounting and Auditing Enforcement Releases) relating to its registrants. It was 44 in 2006.

[2] Average time elapsed between the initiation date and the date the SEC issued AAERs increased by one-third, to 5.6 years in 2006 from 4.1 years in 2001.

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.08.007

overview of existing financial shenanigans and their trend, and (3) suggest a new framework to detect evolutionary financial statement fraud.

The remainder of this paper is organized as follows. We review the application of regression, decision trees, neural networks and Bayesian belief networks in financial statement fraud detection in the next section. We survey the history and trend of contemporary financial fraud in Section 3. We analyze the effectiveness and limitation of existing fraud detection technologies in Sections 4 and 5 and then suggest a framework that addresses the problem when emerging financial fraud is evolutionary. Section 6 concludes the paper with a brief discussion on the insights garnered and possible future research.

2. Review on detection techniques

Classification has been the most popular and the only way used so far to identify fraudulent financial statements [8]. Most financial statement fraud (FSF) auto-detection programs use supervised machine learning methodologies [1,3,4,9–11,14,18,21,22] that usually have a two-stage procedure, where in the first stage a model is trained by using a training sample. The training sample is organized in tuples and attributes, with the class label attribute containing values indicating the pre-defined class to which each tuple belongs. In the second stage, objects are classified through the model obtained from the first stage. After reviewing relevant research in data mining-based financial statement fraud detection literature, we observe that the following five methods have been used so far in this general area. These methods include regression, decision trees, neural networks, Bayesian networks and support vector machines.

Regression is the most widely used method to detect financial statement fraud [1,3,4,11,18,21,22]. Transformations of variables in regression models have also been studied in the context of fraud detection, including logit, stepwise-logistic, multi-criteria decision aid method and exponential generalized beta two. For example, Spathis [21] used a collection of data from 76 firms that include 38 fraudsters and 38 non-fraudulent firms in Greece. They use ten financial variables and logistic multivariate regression to identify the relationship among factors associated with financial statement fraud. A total of ten financial ratios such as the net profit to total assets ratio, the ratio of total debt to total assets, financial distress, the inventories to sales ratio, and the working capital to total assets ratio are selected for examination as potential predictors of FSF. The results indicate that companies with high inventories with respect to sales, high debt to total assets, low net profit to total assets, low working capital to total assets and high financial stress are more likely to manipulate financial statements.

A neural network is another popular data mining technique that has been successfully used to detect financial statement fraud [6,9,14,16,24]. Neural network doesn't assume an attribute's independence and is capable of mining inter-correlated data and is a suitable alternative for problems where some of the assumptions associated with regression are not valid. White [24], nonetheless, has shown that feed-forward neural networks, which require no pre-specified functional form, perform the same stochastic approximation as nonlinear regression. Back propagation neural network allows the network to adapt and has become one of the most popular techniques for prediction and classification problems. The back propagation learning process works in small iterative steps that continuously make small changes to the weights in each neural network layer, which are calculated to reduce the systematic error. The iteration is repeated until the overall error value drops below some pre-determined threshold [13]. Drawbacks of implementing neural networks to discover FSF is that neural network is not accurate if the data is volatile or if the causal functionality evolves in a direction that is not pre-defined.

The objective of decision trees is classification by dividing observations into mutually exclusive and exhaustive subgroups by properly selecting attributes that best separate the sample. Koh and Low [13] construct a decision tree to predict the hidden problems in financial statements by examining the following six variables: quick assets to current liabilities, market value of equity to total assets, total liabilities to total assets, interest payments to earnings before interest and tax, net income to total assets, and retained earnings to total assets.

The above mentioned data mining techniques have generally been shown to be effective in detecting financial statement fraud. However, they are not without limitations. For example, while these techniques are well developed for predictive modeling, they are not as well developed for effect assessment. In particular, test statistics for assessing the effects of independent variables on dependent variables have not yet been constructed for some data mining techniques. Incidentally, this shortcoming also demands with the challenge to develop more effective mechanisms especially in an adaptive economic environment where financial fraudsters learn to circumvent existing automated detection systems.

3. Detection with domain knowledge on FSF

Financial statement fraud, including motivations, opportunities, and rationalizations for management to commit such fraud, has been extensively studied by researchers in finance. Loebbecke et al. [15] suggest a model consisting of three variables that may explain financial statement fraud: (C) the degree to which conditions are such that a financial fraud could be committed, (M) the degree to which the management has a reason or motivation to commit financial fraud, and (A) the degree to which the management has an attitude or set of ethical values such that they would allow themselves to commit management fraud. These three variables together form the assessment model such that the possibility of having financial statement fraud (FSF) can be described as a function.

$$P(FSF) = f(C, M, A) \tag{1}$$

where if C or M or A = 0, then P(FSF) = 0.

According to Rezaee [19], fraud accomplishment can be explained by three variables: (1) conditions, (2) corporate structure and (3) choice. The sufficient incentives and opportunities for a company to commit financial statement fraud can be interpreted by the pattern exhibited by these three variables. We, however, believe that certain combination of the variables mentioned in this model also leads to certain suitable fraud strategies and, by measuring the pattern among the variables and integrating the findings in auto-detection heuristics, we should be able to pinpoint the unique shenanigans and their underlying dynamic to commit fraud.

We first review possible variables that can be utilized in an auto-detection system. In Rezaee's[19] 3C's model as shown in Fig. 1, “conditions” refers to the economic and financial pressures that a corporation faces. Financial pressures, such as pressure to meet analysts' earning estimates, can be a key factor stimulating earnings management and resulting in financial statement fraud. The principle underlying this variable is that financial statement fraud will most probably occur if the benefits for fraudulent management outweigh the associated costs. Management compares the benefit, in terms of possible increase in the company's stock price or the possible savings related to preventing share price from decreasing, with the possible cost of committing fraud in terms of probability and consequences of detection. Financial pressures, such as the inability to meet analysts' earning estimates or decline in quality and quantity of earnings, are often motivations for management commitment in financial frauds.

“Capital structure” refers to the existence of an effective corporate governance mechanism (such as internal control structure and audit committees) that could discourage management from committing